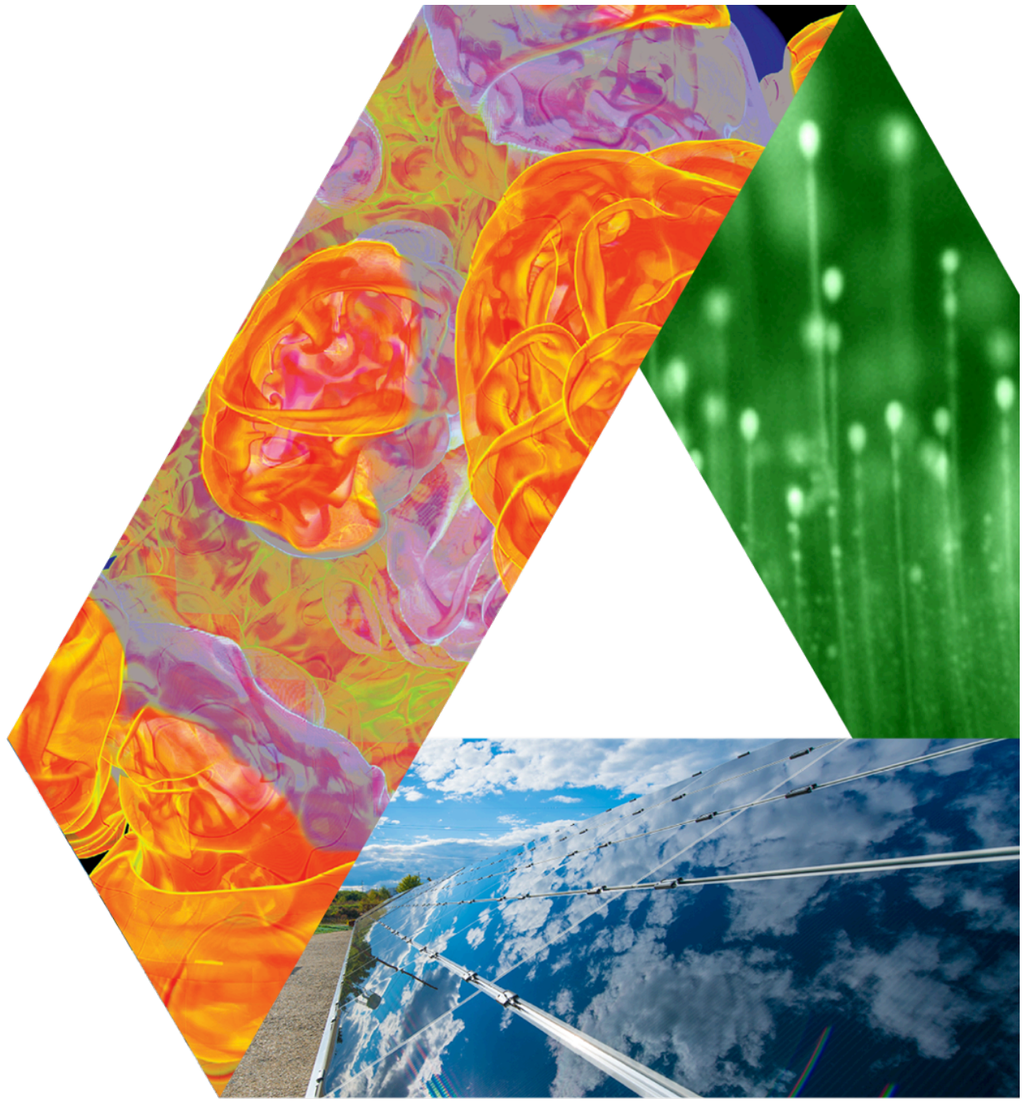
ARGONNE LEADERSHIP COMPUTING FACILITY



Susan Coghlan

ALCF Project Director

Argonne National Laboratory
27 July 2015

DOE LEADERSHIP COMPUTING FACILITY

Leadership Class Resources

- Dedicated to high impact breakthrough open science
- Supported by DOE's Advanced Scientific Computing Research Program
- Two Centers – ALCF & OLCF
- Two of the world's most powerful supercomputers
- Two diverse architectures

ALCF RESOURCES

Mira - compute

- 10 PF IBM BG/Q
- 48K nodes/786K cores
- 786 TB memory
- 5D Torus interconnect
- 26 PB GPFS, 400 GB/s

Cooley – data analytics

- 223 TF
- 126 nodes/1512 Xeon cores/126 Tesla K80 GPUs
- 384 TB (CPU)/3 TB (GPU) memory
- FDR InfiniBand interconnect
- Connected to Mira file systems

Cetus – app d&d

- 840 TF IBM BG/Q
- 4K nodes/64K cores
- 64 TB memory
- 5D Torus interconnect
- Connected to Mira file systems

Vesta – system SW d&d

- 420 TF IBM BG/Q
- 2K nodes/32K cores
- 32 TB memory
- 5D Torus interconnect
- 1 PB GPFS

THREE PRIMARY WAYS TO ACCESS LCF

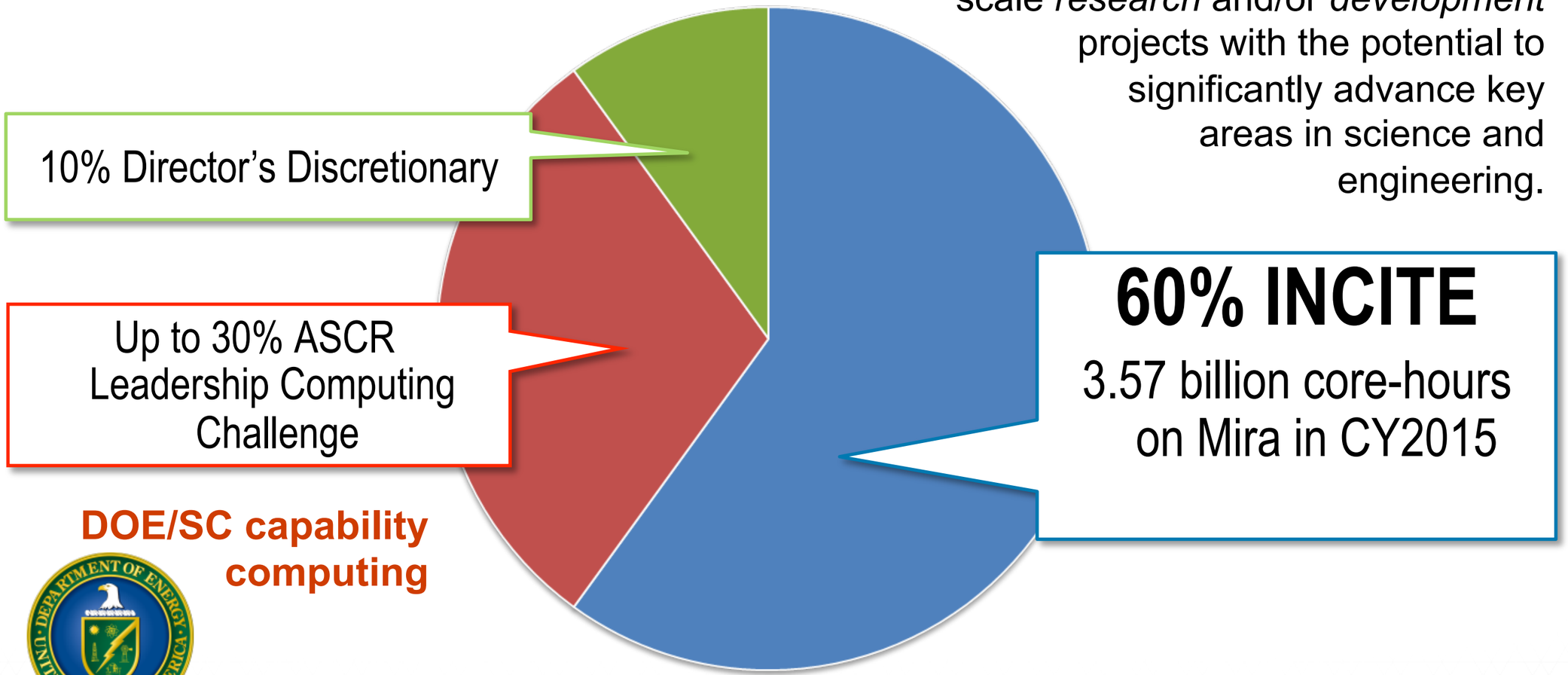
DISTRIBUTION OF ALLOCABLE HOURS

Argonne NATIONAL LABORATORY

OAK RIDGE National Laboratory

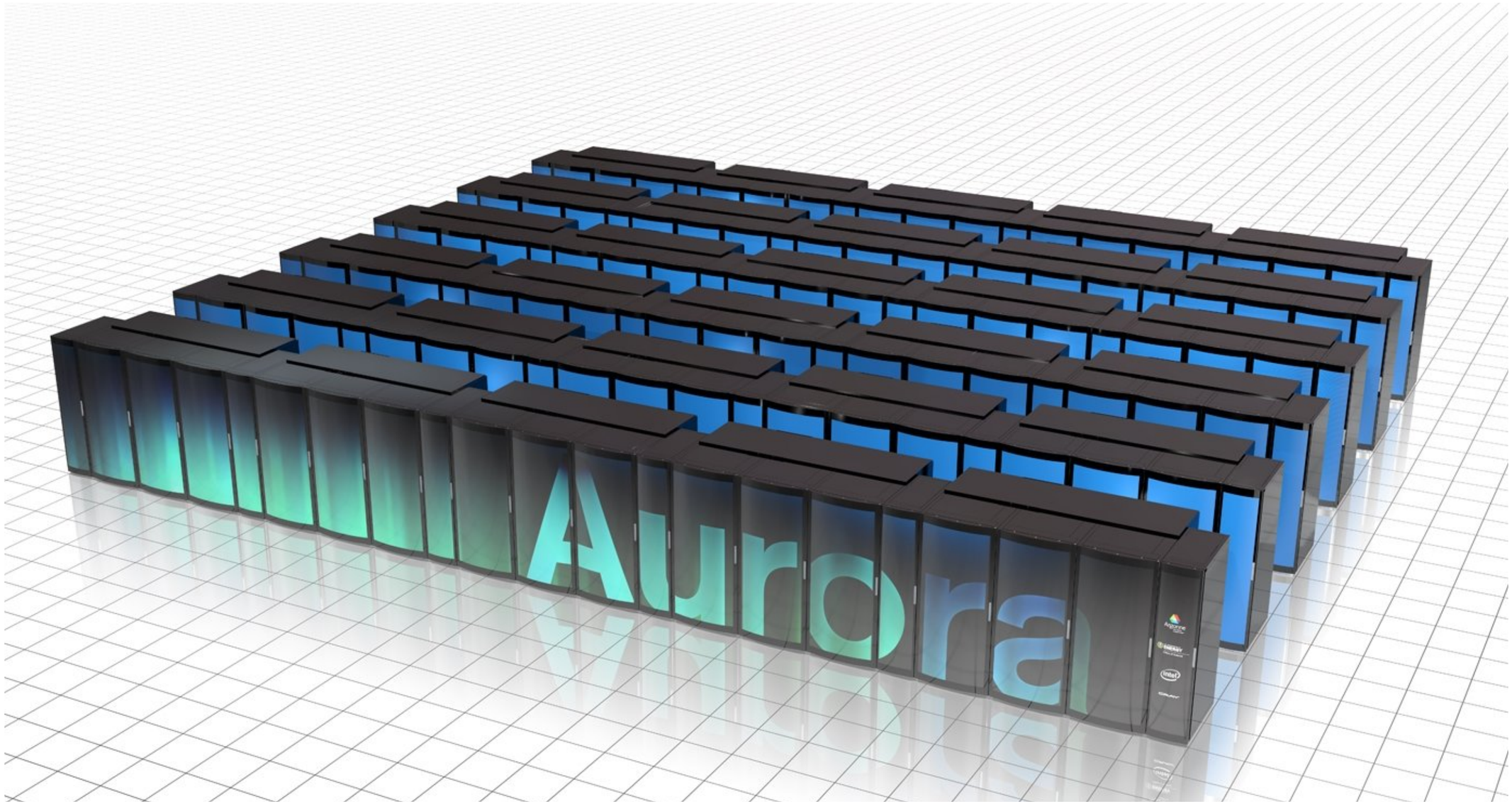
Leadership-class computing

INCITE seeks computationally intensive, large-scale *research* and/or *development* projects with the potential to significantly advance key areas in science and engineering.



DOE/SC capability computing

THE FUTURE



CORAL

COLLABORATION OF OAK RIDGE, ARGONNE, AND LIVERMORE

- Acquire DOE 2018 – 2023 Leadership Computing Capability
- Three leadership class systems – one each at ALCF, LLNL, OLCF
  - With arch diversity between ALCF and OLCF
- ALCF: Intel (Prime) Cray (Integrator)
- OLCF: IBM (Prime)
- LLNL: IBM (Prime)

TWO NEW ALCF SYSTEMS THETA AND AURORA

- Intel Xeon Phi compute architecture
- Deep memory architecture – very fast memory, slower capacity memory, burst buffers

THE RIGHT PATH FOR OUR USERS

- Many core evolution
- Easy to port codes
- Well-balanced between compute, memory, network, and storage
- Robust and well-known Cray user environment combined with Intel innovations

THETA: STEPPING STONE TO AURORA

System Features	Theta Details
Delivery Timeline	2016
Peak System Performance	>8.5 PetaFLOP/s
Compute Node CPU	2nd Generation Intel® Xeon Phi™ processors - **KNL**
Compute Node Count	>2,500 single socket nodes
Compute Platform	Cray XC supercomputing platform
Compute Node Peak Performance	>3 TeraFLOP/s per compute node
Cores Per Node	>60 cores
High Bandwidth On-Package Memory	Up to 16 Gigabytes per compute node
DDR4 Memory	192 Gigabytes per compute node
SSD	128 Gigabytes per compute node
File System	Intel Lustre File System
File System Capacity (Initial)	10 Petabytes
File System throughput (Initial)	200 Gigabytes/s
System Interconnect	Cray Aries Dragonfly topology interconnect
Intel Arch (x86-64) Compatibility	Yes

MIRA -> THETA IMPACT FOR SCIENTISTS

- Same MPI + OpenMP/ pThreads PM
- Dragonfly advantages
  - Much higher connectivity: applications with irregular point-to-point communication will do better
  - Much smaller diameter
- Increased vectorization opportunities
  - Compilers: Intel, Cray, and PGI
  - AVX512: more widely used than QPX, available in other Intel CPUs
  - Wider vector SIMD unit
  - Two independent vector units
- Memory changes
  - Fastest memory is ~12x faster, slower memory is ~2.7x faster
  - Capacity is 13x more per node
- Additional improvements
  - Better single thread performance
  - Larger caches (L1, instruction, L2)
  - Memory per core larger
- Progression of {cores,threads} per node
  - Mira {16,64} → Theta {>60,>240}

AURORA – COMING IN 2018

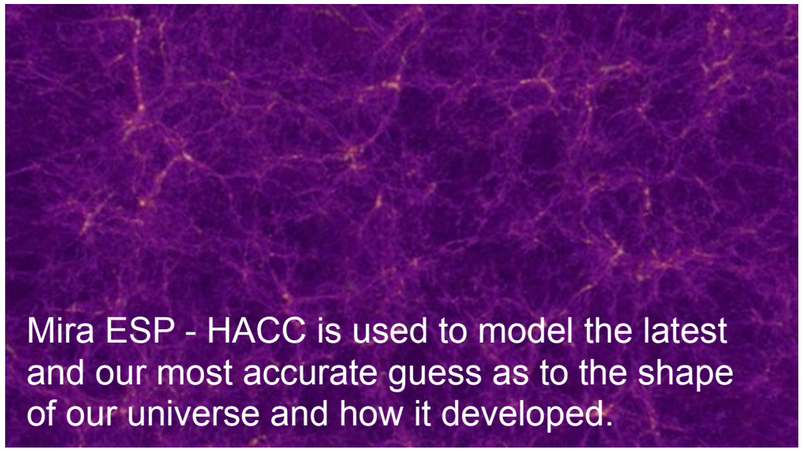
System Feature	Aurora Details
Peak System performance (FLOPs)	180 - 450 PetaFLOPS
Application FOM	13x Mira
Processor	3rd gen Intel Xeon Phi processor - **KNH**
Number of Nodes	>50,000
Compute Platform	Cray Shasta next generation SC platform
HBM, local mem, persistent mem	>7 PetaBytes
High Bandwidth On-Package Memory BW	>30 PB/s
System Interconnect	2nd gen Intel Omni-Path with silicon photonics
Interconnect Aggregate Node Link BW	>2.5 PB/s
Interconnect Bisection BW	>500 TB/s
Interconnect interface	Integrated
Burst Storage Buffer	Intel SSDs
File System	Intel Lustre File System
File System Capacity	>150 PetaBytes
File System Throughput	>1 TeraByte/s
Intel Architecture (x86-64) Compatibility	Yes

THETA -> AURORA IMPACT FOR SCIENTISTS

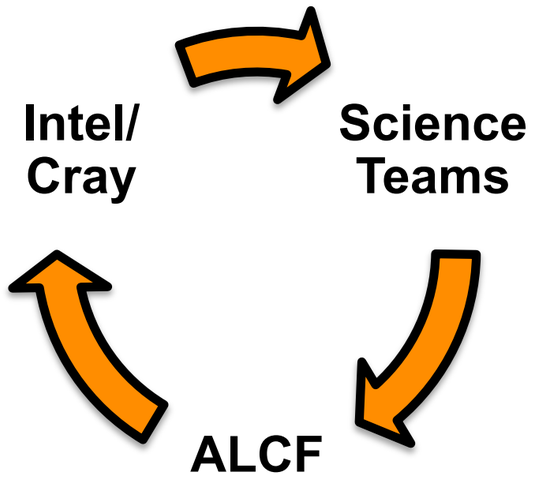
- Next generation Xeon Phi
- Similar number of nodes as Mira
  - Slightly more node concurrency than Theta
- Similar tiered memory and I/O
  - Faster & more HBM, more slower capacity mem
- Interconnect
  - New Intel Omni-Path, but same topology
- Similar Cray software stack
  - +Intel IP improvements
- Same MPI + OpenMP/pThreads PM

ALCF EARLY SCIENCE PROGRAM

GOAL: SCIENCE ON DAY ONE



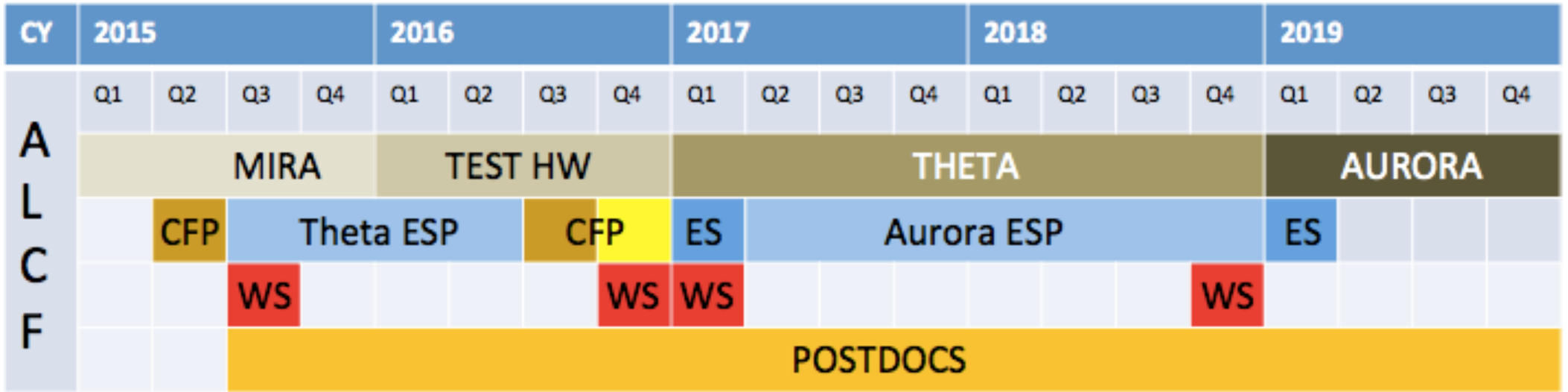
Mira ESP - HACC is used to model the latest and our most accurate guess as to the shape of our universe and how it developed.



- Based on successful Mira ESP pioneering program
- Theta ESP (2015 – 2017)
- Aurora ESP (2016 – 2019)

Call for Aurora ESP proposals 3Q 2016

ALCF ESP CURRENT TIMELINE



Theta	Aurora
1. Apr 2015: ESP Call	1. Jul 2016: ESP call
2. May 2015: ESP Call closed	2. Sep 2016: ESP call closed
3. Jul 2015: ESP teams selected, work begins	3. Dec 2016: ESP projects selected, work begins
4. Aug 2015: ESP kick-off workshop	4. Jan 2017: ESP kick-off workshop
5. Oct 2016: ESP hands-on workshop	5. Nov 2018: ESP hands-on workshop
6. Jan 2017: Early Science dedicated-access period begins	6. Jan 2019: Early Science dedicated-access period begins
7. Mar 2017: Early Science dedicated-access period ends	7. Mar 2019: Early Science dedicated-access period ends

THETA ESP TIER 1 AWARDS

Juan J. Alonso *(Stanford U)* **CFD, renewable energy, engineering** SU2	**Scale-Resolving Simulations of Wind Turbines with SU2** Completion of this project will result in a simulation capability that can be used to design better wind turbines and to layout large wind farms with maximum energy extraction and improved turbine fatigue life. All developments will be available to the community via the SU2 framework.
Fabien Delalondre *(Ecole Federale Polytechnique de Lausanne (EPFL))* **Neuroscience, biology** CoreNeuron	**Large Scale Simulation of Brain Tissue: Blue Brain Project** While much is known about the brain, little is understood. Integration through data-driven brain tissue models is a novel method in the scientific repertoire. The project will advance simulation tools, widen the ability for scientific questions on brain plasticity, and trigger valuable discussions with the HPC community.
Giulia Galli *(U Chicago, Argonne)* **Materials Science, Chemistry** Qbox, West	**First-Principles Simulations of Functional Materials for Energy Conversion** Properties of materials to be used for solar and thermal energy conversion will be optimized at an unprecedented level of accuracy—by combining ab initio molecular dynamics and post-density functional theory methods—thus providing truly predictive tools, ultimately for device performance, within a MGI material design framework.

THETA ESP TIER 1 AWARDS

Katrin Heitmann *(Argonne)* **Cosmology, high energy physics** HACC	**Next-Generation Cosmology Simulations with HACC: Challenges from Baryons** By confronting new observations with sophisticated simulations the project will further the understanding of astrophysical processes on small scales. At the same time they will disentangle these processes from fundamental physics and therefore help mitigate one of the major sources of systematic uncertainties for upcoming cosmological surveys.
Alexei Khokhlov *(U Chicago)* **Combustion, CFD** HSCD	**Direct numerical simulations of flame propagation in hydrogen-oxygen mixtures in closed vessels** First-principles understanding and quantitative prediction of flame acceleration and DDT in hydrogen is important for industrial and public safety of hydrogen fuels, and for safety of certain types of water-cooled nuclear reactors.
Benoit Roux *(U Chicago, Argonne)* **Biophysics, molecular dynamics** NAMD	**Free Energy Landscapes of Membrane Transport Proteins** An atomistic picture of membrane transport proteins is a critical component of understanding of a broad range of biological functions. This work will utilize computational models to provide both detailed visualizations of large protein motions as well as quantitative predictions into the energetics of these processes.

THETA ESP TIER 2 AWARDS

Volker Blum (Duke U) **Materials science** FHI-aims/Gator	**Electronic Structure Based Discovery of Hybrid Photovoltaic Materials on Next-Generation HPC Platforms**
Christos Frouzakis (ETHZ) **Combustion, CFD** Nek5000	**Flow, mixing and combustion of transient turbulent gaseous jets in confined cylindrical geometries**
Mark Gordon (Iowa State U) **Chemistry** GAMESS	**Advanced Electronic Structure Methods for Heterogeneous Catalysis and Separation of Heavy Metals**
Kenneth Jansen (U of Colorado at Boulder) **CFD, aerodynamics, nuclear energy** PHASTA	**Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control**
Paul Mackenzie (Fermilab) **Lattice QCD, high energy physics** MILC/CPS	**The Hadronic Contribution to the Anomalous Magnetic Moment of the Muon**
Steven Pieper (Argonne) **Nuclear physics** GFMC	**Quantum Monte Carlo Calculations in Nuclear Theory**

TOOLS AND LIBRARIES COLLABORATION

Debuggers
- Allinea *(C. January)*
- RogueWave *(J. DelSignore, C. Schneider, S. Lawrence)*

Performance Tools
- HPCToolkit *(J. Mellor-Crummey, Rice U.)*
- TAU *(S. Shende, ParaTools)*
- PAPI *(H. McCraw, UTK)*

Compilers
- *LLVM (Hal Finkel, ALCF)*

OS
- Argo *(K. Iskra, K. Yoshii, Argonne)*

Libraries
- PetSC *(B. Smith, Argonne)*
- Elementall *(J. Poulson, Stanford)*
- Intel MKL ScaLAPACK *(Intel)*
- LIBXSMM *(Intel)*
- ELPA *(Intel)*
- NWChem packages *(Intel)*

Programming Models
- MPICH *(P. Balaji, Argonne)*
- BerkelyUPC *(P. Hargrove, Y. Zheng, LBNL)*
- ARMCI-MPI *(Intel)*
- CommAgent MPI *(Intel)*
- EP-lib *(Intel)*

I/O
- GLEAN *(V Vishwanath, ANL)*
- HDF5 *(Q. Koziol , M. Chaarawi, J. Soumagne, S. Breitenfeld, N. Fortner, UIUC)*
- Mercury *(R. Ross, Argonne)*
- MPI *(R. Latham, Argonne)*
- Darshan *(P. Carns, Argonne)*
- Parallel I/O library (PIO) *(J. Edwards, J.Dennis, NCAR; J. Krishna, ANL)*

Other
- Model Coupling Toolkit *(R. Jacob, Argonne)*
- Common Infrastructure for Modeling the Earth (CIME) *(M. Vertenstein, J. Edwards, NCAR; R. Jacob, Argonne)*

- Porting for Performance: software & programming environment collaboration to enable high performance computational science

ARCHITECTURE AND PERFORMANCE PORTABILITY

Application portability among ALCF, OLCF, and NERSC architectures is critical concern of ASCR

- Application developers target wide range of architectures
- Maintaining multiple code versions is difficult
- Porting to different architectures is time-consuming
- Many Principal Investigators have allocations on multiple resources
- Applications far outlive any computer system

Improve data locality and thread parallelism

- Many-core or GPU optimizations improve performance on all architectures
- Exposed fine grain parallelism transitions more easily between architectures
- Data locality optimized code design also improves portability

Use portable libraries

- Library developers deal with portability challenges
- Many libraries are DOE supported

MPI+OpenMP 4.0 could emerge as common programming model

- Significant work is still necessary
- All ASCR centers are on the OpenMP standards committee

Encourage portable and flexible software development

- Use open and portable programming models
- Avoid architecture specific models such as Intel TBB, NVIDIA CUDA
- Use good coding practices: parameterized threading, flexible data structure allocation, task load balancing, etc.

Argonne Leadership Computing Facility

Our people set us apart

Argonne
NATIONAL
LABORATORY